Secret Contracts for Efficient Partnerships*

David Rahman† and Ichiro Obara‡

August 30, 2007

Abstract

By allocating different information to team members, secret contracts can provide better incentives to perform with an intuitive organizational design. For instance, they may help to monitor monitors, and attain approximately efficient partnerships by appointing a secret principal. More generally, secret contracts highlight a rich duality between enforceability and identifiability. It naturally yields necessary and sufficient conditions on a monitoring technology for any team using linear transfers to approximate efficiency (with and without budget balance). The duality is far-reaching: it is robust to complications in the basic model such as environments with limited liability and participation constraints.

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

*This paper subsumes the previously circulated Rahman (2005a) and Obara and Rahman (2006). Many thanks are owed to Harold Demsetz, Larry Jones, Narayana Kocherlakota, David Levine, Roger Myerson, Joe Ostroy, Phil Reny, Joel Sobel, Bill Zame, the co-editor Larry Samuelson and three anonymous referees for help with previous drafts. We are also grateful to seminar audiences at UCLA, UC San Diego, Caltech, University of Chicago, Universidad Carlos III de Madrid, Universitat Pompeu Fabra, University of Minnesota, University of Arizona, and Università Bocconi.

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

‡Department of Economics, University of California, Los Angeles.

Contents

But who will monitor the monitor? Alchian and Demsetz (1972, p. 782)

1 Introduction

Ann owns a restaurant. She hires Bob to tally up the till every night and report back any mismatch between the till and that night's bills. Ann can motivate Bob to exert such effort and report truthfully any mismatch by secretly taking some money from the till herself with positive probability and offering him the following incentive scheme: if Ann took some money, she will pay Bob his wage only when he reports a mismatch; if Ann did not take any money, she will pay Bob only when a mismatch is not reported.

Bob faces a secret contract: his report-contingent wage is unknown to him a priori (it depends on whether or not Ann secretly took some money). If Bob fails to exert effort, he won't know what to report in order to secure his wage. However, if he does his job he'll discover whether or not there is a mismatch and deduce from this Ann's behavior. Only after tallying the till will Bob know what to report in order to receive his wage, which turns out to be optimally truthful.

This paper studies contracts like Bob's and how they might help organizations to function productively. By allocating different information to team members, secret contracts often provide better incentives to perform with an intuitive organizational design. Thus, they give Bob incentives to acquire costly information and reveal it. In general, they provide a way of "monitoring the monitor" (Section 2.1), and can yield approximately efficient partnerships by appointing a "secret principal" (Section 2.2).

A rich duality between enforceability and identifiability—more specifically, between incentive compatible contracts and indistinguishable deviation plans—is exploited. It leads us to identify teams that can approximate efficiency (with and without budget-balanced transfers) by means of their "monitoring technology" (Section 3). This duality is far-reaching: it is amenable to complications in the basic model such as individual rationality and limited liability (Section 4).

1.1 Monitors and Principals

According to Alchian and Demsetz (1972, p. 778, their footnote), [*t*]*wo key demands are placed on an economic organization—metering input productivity and metering rewards.*[1] At the heart of their "metering problem" lies the question of how to give incentives to monitors, which they answered by making the monitor residual claimant. However, this can leave the monitor with incentives to misreport input productivity if his report influences input rewards, like workers' wages, since—given efforts—paying workers hurts him directly.[2]

On the other hand, Holmström (1982, p. 325) argues that ... *the principal's role is not essentially one of monitoring ... the principal's primary role is to break the budget-balance constraint*. Where Alchian and Demsetz seem to overemphasize the role of monitoring in organizations, Holmström seems to underemphasize it. He provides incentives with "team punishments" that reward all agents when output is good and punish them all when it is bad. Assuming that output is publicly verifiable, he finds little role for monitoring,[3] and perhaps as a result Holmström (1982, p. 339) concludes wondering: ... *how should output be shared so as to provide all members of the organization* (*including monitors*) *with the best incentives to perform?*

Secret contracts motivate monitors: If the principal secretly recommends a worker to shirk or work, both with some probability (the worker can easily be motivated to willingly obey recommendations), and pays the monitor only if he reports back the recommendation, then—like Bob—the monitor will prefer to exert effort and report truthfully. To implement such contracts, the team requires (i) a disinterested mediator or machine that makes confidential, verifiable but non-binding recommendations to players, and (ii) transfers that depend on the mediator's recommendation as well as the monitor's report. As this requirement suggests, incentive compatibility of secret contracts is described here by Myerson's (1986) *communication equilibrium*.

[1] *Meter means to measure and also to apportion. One can meter* (*measure*) *output and one can also meter* (*control*) *the output. We use the word to denote both; the context should indicate which.*

[2] A comparable argument was put forward by Strausz (1997) by observing that delegated monitoring dominates monitoring by a principal who cannot commit to his agent that he will verify the agent's effort when it is only privately verifiable. However, Strausz assumes that monitoring signals are "hard evidence," so a monitor cannot misreport his information.

[3] Intuitively, if output were not publicly verifiable then his team punishments would no longer provide the right incentives: monitors would always report good output to secure payment and shirk from their monitoring responsibilities to save on effort. Knowing this, workers would also shirk.

Monitoring adds value only insofar as it helps to provide incentives. Heuristically, if monitors never monitor then workers will not work, so costly monitoring may be worthwhile. Nevertheless, it is cost-efficient to do so as little as necessary. This leads naturally to *approximate efficiency* as the appropriate optimality criterion for a team with costly monitoring, especially when having access to *linear transfers*. For example, secret (mixed) monitoring of workers with small but positive probability together with large punishments if caught shirking saves costs while providing incentives.

This use of mixed strategies to approximate efficiency was developed by Legros and Matthews (1993) in Nash equilibrium with public, deterministic output. Not only can secret contracts exploit such mixing, too, but also (and in addition to monitoring the monitor) they can improve a team's contractual prospects even in the restricted setting of publicly verifiable output, as the secret principal demonstrates.

To see this, recall the partnership problem of Radner et al. (1986). It shows that no budget-balanced linear transfers contingent only on output can approximate efficiency in a team whose members can either work or shirk and whose joint output is (publicly verifiable and) either high or low with a probability that is increasing only in the number of workers. A secret principal approximates efficiency: With arbitrarily large probability, suppose everyone is recommended to work, and paid nothing regardless. With complementary probability, everybody is told to work except for one randomly picked team member, who is secretly told to shirk. This individual must pay everyone else if output is high and be paid by everyone else if output is low. Such a scheme is incentive compatible with large payments, budget-balanced, approximately efficient.

1.2 Enforceability and Identifiability

Assuming correlated equilibrium and approximate efficiency/enforceability renders linear our formal description of incentive compatible contracts. In other words, some given team behavior is approximately implementable with incentive compatible secret transfers if and only if a certain family of linear inequalities is satisfied. A duality theory of contracts therefore obtains as a result of this linearity, with basic implications for understanding incentives. We take advantage of this duality throughout the paper, which prevails over gradual complications in our basic model. Technically, our linear methods rely on Rahman (2005a) to extend those of Nau and McCardle (1990) and d'Aspremont and Gérard-Varet (1998) with substantially stronger results.

Duality yields two sides of the same coin, two opposite views of the same problem—in our case, a metering problem. As the title of this subsection—taken from Fudenberg et al. (1994, p. 1013)—suggests, enforceable contracts and unidentifiable deviation plans are mutually dual variables. As such, two natural descriptions of a team's monitoring technology emerge from each point of view. The primal side of the coin describes when contracts are approximately enforceable, whereas the dual side describes when deviation plans cannot be identified. Thus, the smaller the set of unidentifiable deviation plans, the larger the set of enforceable contracts—like a cone and its polar. In the limit, our main results (Theorems 1 and 3) provide intuitive conditions on a monitoring technology that are necessary and sufficient for any team outcome to be approximately enforceable via secret contracts (with and without budget balance).

Theorem 1 provides a minimal requirement on a team's monitoring technology, called detecting unilateral disobedience (DUD), that characterizes approximate enforceability with secret contracts of any team outcome. Intuitively, for every disobedient deviation plan there must be some correlated strategy (not necessarily the same one for every plan) that renders the disobedience statistically detectable. (Dishonesty may remain undetectable, though.) DUD turns out to be weak and generic.[4]

Restricting attention to budget-balanced secret contracts, Theorem 3 characterizes approximate enforceability of team behavior with a stronger condition, called identifying obedient players (IOP). Intuitively, IOP requires that—in addition to DUD—it is possible to statistically identify some player as obedient upon any disobedience. IOP is weak[5] and generic,[6] too. Intuitively, IOP delivers incentives with budget balance by rewarding those known to be "innocent" while punishing all others.

Our use of duality facilitates the study of other restrictions to the metering problem, like limited liability and individual rationality. Well-known results, such as that only total liability matters when providing a team with incentives or that reasonably low participation constraints don't bind even with budget balance, are extended to this framework without complications. Exact implementation fits relatively nicely, too.

[4]DUD is weaker than comparable conditions in Compte (1998) and Obara (2005). Restricted to public monitoring, DUD is weaker than local individual full rank, of d'Aspremont and Gérard-Varet (1998), which in turn is weaker than the condition in Legros and Matsushima (1991).

[5]IOP is weaker than comparable conditions such as those in Kandori and Matsushima (1998), Aoyagi (2005), and Tomala (2005). With public monitoring, it is still weaker than the compatibility of d'Aspremont and Gérard-Varet (1998) and even Kandori's (2003) version of pairwise full rank.

[6]Like DUD, IOP is "as generic if not more" than other conditions in the literature (Section **??**).

Further discussion of secret contracts, particularly as regards the theory of mechanism design and their susceptibility to collusion, is deferred to the conclusion (Section 5).

2 Examples

We begin our formal analysis of secret contracts with two important, motivating examples mentioned in the introduction: monitoring the monitor, and the secret principal. The first example studies an environment involving contractual variations on a three-player game that attempts to typify the strategic interaction between a principal, an agent, and a monitor. The second example finds an intuitive way of attaining approximately efficient partnership with budget-balanced contracts.

2.1 Robinson and Friday

There are three players. The first is Robinson, who can either monitor or shirk. The second is Friday, who can either work or shirk. The third player is a so-called mediating principal, a disinterested party who makes recommendations and enforces contingent contractual payments. For simplicity, suppose the principal's utility is constant regardless of the outcome of the game. Robinson (the row player) and Friday (the column player) interact according to the left bi-matrix below.

	work	shirk
monitor	$2,-1$	$-1,0$
shirk	$3,-1$	$0,0$

Utility Payoffs

	work	shirk
monitor	$1,0$	$0,1$
shirk	$1/2,1/2$	$1/2,1/2$

Signal Probabilities

The action profile (shirk,work) is Pareto efficient, since Robinson finds monitoring costly and it does not intrinsically add value. However, this strategy profile is not incentive compatible by itself, since Friday always prefers to shirk rather than work.

The team's monitoring technology is given by a set $S = \{g, b\}$—so there are only two possible signals contingent upon which contracts may be written—together with the conditional probability system given by the right bi-matrix above. In words, if Robinson shirks then both signals are equiprobable, whereas if he monitors then the realized signal will accurately identify whether or not Friday worked. Contractual payments are assumed to be denominated in a private good ("money") that enters players' utility linearly with unit marginal utility.

Clearly, the efficient strategy profile (shirk,work) cannot be implemented.[7] However, we can get arbitrarily close: When signals are publicly verifiable, the correlated strategy[8] $\sigma[(\text{monitor,work})] + (1-\sigma)[(\text{shirk,work})]$ can be implemented for any $\sigma \in (0,1]$ with Holmström's *team punishments*. For example, paying Robinson \$2 and Friday $\$1/\sigma$ if g and both players zero if b makes (shirk,work) approximately implementable.

If only Robinson observes the signal, and it is not verifiable, then for the principal to write signal-contingent contracts, he must first solicit the realizations from Robinson, who may in principle misreport them. Notice that now team punishments break down, since not only will Robinson always report g and shirk, but also Friday will shirk. Furthermore, if Robinson was rewarded independently of his report then although he would happily tell the truth, he would find no reason to monitor.

Another possibility is to have Friday mix between working and shirking. On its own, this strategy doesn't change Robinson's incentives to either lie or shirk. However, if the principal and Friday correlate their play without Robinson knowing when, it is possible to "cross-check" Robinson's report, thereby "monitoring the monitor."

Specifically, the following correlated strategy is incentive compatible given $\mu \in (0,1)$: (i) Robinson is told to monitor with probability σ (and shirk with probability $1-\sigma$), (ii) Friday is independently told to work with probability μ (to shirk with $1-\mu$), and (iii) the principal correlates his contractual strategy with players' recommendations:

	(monitor,work)	(monitor,shirk)	(shirk,work)	(shirk,shirk)
g	$1/\mu, 1/\sigma$	$0,0$	$0,0$	$0,0$
b	$0,0$	$1/(1-\mu),0$	$0,0$	$0,0$

The numbers on the left are Robinson's contingent payments, and those on the right are Friday's. Thus, Robinson is paid $\$1/\mu$ if he reports g when (monitor,work) was recommended and $\$1/(1-\mu)$ if he reports b when (monitor,shirk) was recommended. It is easily seen that honesty and obedience to the mediator is incentive compatible. This contract approximately implements (shirk,work) by letting $\sigma \to 0$ and $\mu \to 1$.

[7] If Robinson shirks then no signal-contingent contract can compensate Friday more when working than shirking, since each signal carries the same probability regardless of Friday's effort.

[8] As a matter of notation, let $[a]$ stand for Dirac measure (or the pure strategy profile a living in the space of correlated strategies) for any action profile a.

Particularly distinguishing properties of this contract are that Robinson does not directly observe the principal's recommendation to Friday, and that Robinson has the incentive to monitor inasmuch as he is rewarded for reporting accuracy. Notice also that Robinson's report only confirms to the principal his recommendation to Friday. As such, the principal strips away Robinson's a priori informational advantage, which is why his surplus can be extracted. The principal allocates private information to approximate efficiency, so a team without asymmetric information may prefer to create some as part of its organizational design. A salient problem of the contract is not being robust to "collusion:" If Friday told Robinson his recommendation then both players could save on effort. We do not address collusion formally in this paper, but see Section 5.3 for a way to dissuade extra-contractual communication. On the other hand, there is no other way for Friday to work with positive probability—not without secrets. Finally, it is impossible to approximate efficiency with budget balance, but a reasonably different monitoring technology permits budget balanced approximate efficiency (Example 3) only with secret contracts, robust to this collusion.

2.2 Secret Principal

A team has n individuals. Each team member 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.

Radner et al. (1986) showed that in this environment there do not exist budget-balanced output-contingent linear transfers to induce everyone to work, not even approximately. One arrangement that is not approximately efficient but nevertheless induces most people to work is appointing Holmström's principal. Call this player 1 and define transfers as follows. For $i = 2, \ldots, n$ let $\zeta_i(g) = k$ and $\zeta_i(b) = 0$ be player i's output-contingent linear transfer, for some $k \geq 0$. Let player 1's transfer equal

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

By construction, the budget is balanced. Everyone but player 1 will work if k is sufficiently large. However, player 1 has the incentive to shirk. This contract follows Holmström's suggestion to the letter: Player 1 is a "fixed" principal who absorbs the incentive payments to all others by "breaking" everyone else's budget constraint.

Allowing now for secret contracts, consider the following scheme. For any small $\varepsilon > 0$, a mediator asks every individual to work (call this event $\mathbf{1}$) with probability $1 - \varepsilon$. With probability ε, he picks some player i at random (with probability ε/n for all i) and asks him secretly to shirk, while telling all others 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 player i's contingent transfer if the mediator asked everyone to work. Otherwise, if player i was secretly told to shirk, for $j \neq i$ let $\zeta_j(g|\mathbf{1}_{-i}) = k$ and $\zeta_j(b|\mathbf{1}_{-i}) = 0$ be player j's transfer. For player i, let

$$\zeta_i = -\sum_{j \neq i} \zeta_j.$$

Clearly, this contract is budget-balanced. It is also incentive compatible. Indeed, if player i is recommended to work, incentive compatibility requires that

$$\frac{\varepsilon(n-1)}{n} P(n-1)k - c \;\geq\; \frac{\varepsilon(n-1)}{n} P(n-2)k,$$

which is satisfied if k is sufficiently large. If player i is asked to shirk, we require

$$-(n-1)P(n-1)k \;\geq\; -(n-1)P(n)k - c,$$

which always holds.

Therefore, this contract implements the efficient outcome with probability $1 - \varepsilon$ and a slightly inefficient outcome with probability ε. Since ε can be made arbitrarily small (by choosing an appropriate reward k), we obtain an approximately efficient partnership. The role of principal is not fixed here. It is randomly assigned with very small probability to make negligible the loss from having a principal.

3 Model

Let $I = \{1, \ldots, n\}$ be a finite set of players, A_i a finite set of actions available to player $i \in I$, and $A = \prod_i A_i$ the (nonempty) space of action profiles. Actions are neither verifiable nor directly observable. A *correlated strategy* is a probability measure $\sigma \in \Delta(A)$. Let $v_i(a)$ be the utility to player $i \in I$ from action profile $a \in A$.

Let S_i be a finite set of *private signals* observable only by individual member $i \in I$ and S_0 a finite set of *publicly verifiable* signals. Let

$$S := \prod_{j=0}^{n} S_j$$

be the (nonempty) product space of all observable signals. A *monitoring technology* is a measure-valued map $\Pr : A \to \Delta(S)$, where $\Pr(s|a)$ stands for the conditional probability that $s = (s_0, s_1, \ldots, s_n) \in S$ was observed given that the team played $a = (a_1, \ldots, a_n) \in A$. For every $s \in S$, suppose $\Pr(s|a) > 0$ for some $a \in A$.

Assume that the team has access to *linear transfers*. An *incentive scheme* is any map $\zeta : I \times A \times S \to \mathbb{R}$ that assigns monetary transfers contingent on individuals, recommended actions, and *reported* signals. It is assumed that recommendations are verifiable.[9] Rather than focus on incentive schemes ζ, we will also study *probability weighted* transfers, $\xi : I \times A \times S \to \mathbb{R}$. For any recommendation $a \in A$ with $\sigma(a) > 0$, one may think of ξ as solving $\xi_i(a, s) = \sigma(a)\zeta_i(a, s)$ for some ζ. For any $a \in A$ with $\sigma(a) = 0$ and $\xi(a) \neq 0$, one may think of ξ as either arising from unbounded incentive schemes (i.e., $\zeta_i(a, s) = \pm\infty$) or as the limit of a sequence $\{\sigma^m \zeta^m\}$. This change of variables from ζ to ξ is explained further in Section 4.1.

The timing of team members' interaction runs as follows. Firstly, players agree upon some *contract* (σ, ζ) consisting of a correlated strategy σ and an incentive scheme ζ. A profile of recommendations is drawn according to σ and made to players confidentially and verifiably by some machine. Players then simultaneously take some action. Afterwards, they observe their private signals and submit a verifiable report of their observations (given by an element of their personal signal space) before observing the public signal (not essential, just simplifying). Finally, recommendation- and report-contingent transfers are made according to ζ.

[9] This assumption is without loss of generality: If recommendations were not directly verifiable, then players could be asked to announce theirs as verifiable messages.

If every player obeys his recommendation and reports truthfully, the expected utility to player i (before recommendations are actually made) from a contract (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, Mr. 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 Mr. i privately observes s_i. Let R_i be the set of all reporting strategies for player i. The *truthful reporting strategy* is the identity map $\tau_i : S_i \to S_i$ with $\tau_i(s_i) = s_i$. Thus, both $\zeta_i(a, s_{-i}, \tau_i(s_i)) = \zeta_i(a,s)$ and $\xi_i(a, s_{-i}, \tau_i(s_i)) = \xi_i(a,s)$.[10] The space of pure *deviations* for i is therefore $A_i \times R_i$.

For every player i and every deviation (b_i, ρ_i), the conditional probability that signal profile s will be reported when everyone else is honest and plays $a_{-i} \in A_{-i}$ equals

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

When all other players are honest and obedient, the utility to i from deviating to (b_i, ρ_i) conditional on being recommended to play a_i under contract (σ, ζ) equals

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

where $\sigma(a_i) = \sum_{a_{-i}} \sigma(a) > 0$ is the probability that a_i was recommended.

A team's *metering problem* is to find a contract (σ, ζ) that makes incentive compatible obeying recommended behavior as well as honest reporting of monitoring signals. This is captured by the following family of inequalities.

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

The left-hand side reflects the *deviation gain* in terms of utility[11] for a player i from playing b_i when asked to play a_i. The right-hand side reflects his *contractual loss* from deviating to (b_i, ρ_i) relative to honesty and obedience (i.e., playing a_i after being

[10] We will often use the notation $s = (s_{-i}, s_i)$ and $a = (a_{-i}, a_i)$ for any i, where $s_i \in S_i$ and $s_{-i} \in S_{-i} = \prod_{j \neq i} S_j$; similarly for A_{-i}.

[11] Specifically, in terms of probability weighted utility, weighted by $\sigma(a_i)$. If a_i is never recommended then $\sigma(a_i) = 0$ and both sides of the inequality equal zero.

asked to do so and reporting according to τ_i). 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 1. A correlated strategy σ is *exactly enforceable* (or simply enforceable) if there exists an incentive scheme $\zeta : I \times A \times S \to \mathbb{R}$ to satisfy (∗) for all (i, a_i, b_i, ρ_i). Call σ *exactly enforceable with budget balance* if it is exactly enforceable and

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

A correlated strategy σ is *approximately enforceable* if there exists a sequence of contracts $\{(\sigma^m, \zeta^m)\}$ such that (σ^m, ζ^m) satisfies (∗) for every $m \in \mathbb{N}$ and $\sigma^m \to \sigma$. Call σ *approximately enforceable with budget balance* if it is approximately enforceable and ζ^m satisfies (∗∗) for all m.

A correlated strategy is approximately enforceable if it is the limit of exactly enforceable ones. Approximate enforcement with budget balance requires that the budget be balanced along the way, not just asymptotically. For example, in Robinson and Friday (Section 2.1) the correlated strategy [(shirk,work)] is approximately enforceable but not enforceable. In the secret principal (Section 2.2), everybody working is approximately enforceable with budget balance, but not exactly enforceable with budget balance, although it is exactly enforceable (without budget balance).

Before solving the model, a little more notation will be useful. A *deviation plan* for any player 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. Given $\sigma \in \Delta(A)$, let $\Pr(\sigma) \in \mathbb{R}^S$ be the vector defined by $\Pr(\sigma)(s) = \sum_a \sigma(a) \Pr(s|a)$. Intuitively, $\Pr(\sigma)$ is the expected vector of report probabilities if everyone is honest and obediently playing according to σ. Let $\Pr(\sigma, \alpha_i) \in \mathbb{R}^S$, defined pointwise by

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

be the expected vector of probabilities if player i deviates from σ according to α_i.

A deviation plan α_i is *disobedient* if $\alpha_i(\rho_i, b_i|a_i) > 0$ for some $a_i \neq b_i$, i.e., it disobeys some recommendation a_i with positive probability. A disobedient deviation plan may be "honest," i.e., ρ_i may equal τ_i with probability one after every recommendation. Although dishonesty is arguably a form of disobedience, it will be useful in the sequel to distinguish between them.

3.1 Detection

Definition 2 (Detection). A deviation plan α_i for player i is called *undetectable* if

$$\forall \sigma \in \Delta(A), \qquad \Pr(\sigma) \;=\; \Pr(\sigma, \alpha_i).$$

Call α_i *detectable* if it is not undetectable, i.e., $\Pr(\sigma) \neq \Pr(\sigma, \alpha_i)$ for some $\sigma \in \Delta(A)$.

Intuitively, a deviation plan α_i is undetectable if the probability of reported signals induced by α_i, $\Pr(\sigma, \alpha_i)$, coincides with that arising from honesty and obedience, $\Pr(\sigma)$, *regardless of the team's correlated strategy*, σ, assuming that others are honest and obedient. Detectability is a weak requirement. Undetectable deviation plans may be defined equivalently by $\Pr(a) = \Pr(a, \alpha_i)$ for every $a \in A$ due to linearity, but it seems to be a less intuitive description of detectability, albeit more tractable.[12]

Definition 3 (DUD). A monitoring technology Pr *detects unilateral disobedience* (DUD) if every disobedient deviation plan is detectable.

DUD is intuitively defined. Formally, note that different correlated strategies may be used to decide whether or not different disobedient deviation plans are detectable. This is one important aspect that renders DUD substantially weaker than other conditions in the literature. A detailed comparison will be made shortly, but first let us characterize DUD in terms of approximate enforceability.

Definition 4 (PSI). A monitoring technology Pr *provides strict incentives* (PSI) if there exists a probability weighted incentive scheme $\xi : I \times A \times S \to \mathbb{R}$ such that

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

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

If the left-hand side above is interpreted as a player's *deviation gain* from playing b_i when recommended to play a_i, then PSI implies that for any given deviation gains by the players, there is an incentive scheme such that any deviator's contractual loss outweighs his deviation gain after every recommendation. It may appear that PSI is a rather strong condition on a monitoring technology, in contrast with the weakness of DUD argued below (Example 1). As it turns out, both conditions are equivalent, in fact mutually dual.

[12] For a differently tractable version of DUD (without using reporting strategies), see Lemma B.1.

Theorem 1. *A monitoring technology detects unilateral disobedience if and only if it provides strict incentives.*

Proof. By the Alternative Theorem (Rockafellar, 1970, Theorem 22.2, p. 198), a given monitoring technology Pr fails to provide strict incentives if and only if there exists a vector $\lambda \geq 0$ such that $\lambda_i(a_i, b_i, \rho_i) > 0$ for some (i, a_i, b_i, ρ_i) with $a_i \neq b_i$ and

$$\forall (a, s), \qquad \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i)(\Pr(s|a) - \Pr(s|a_{-i}, b_i, \rho_i)) \;=\; 0.$$

Let α_i be the deviation plan defined pointwise by

$$\alpha_i(b_i, \rho_i|a_i) := \begin{cases} \lambda_i(a_i, b_i, \rho_i)/\sum_{(b_i', \rho_i')} \lambda_i(a_i, b_i', \rho_i') & \text{if } \sum_{(b_i', \rho_i')} \lambda_i(a_i, b_i', \rho_i') > 0, \text{ and} \\ [(a_i, \tau_i)]\,(b_i, \rho_i) & \text{otherwise (where } [\cdot] \text{ denotes Dirac measure).} \end{cases}$$

By construction, α_i is disobedient and undetectable: DUD fails. □

This proof describes the duality between identifiability and enforceability via secret contracts. The next result, which may be viewed as a corollary, characterizes DUD as the weakest identifiability required for any action to be approximately enforceable.

Corollary 1. *A monitoring technology detects unilateral disobedience if and only if any team with any profile of utility functions can approximately enforce any correlated strategy with secret contracts.*

Corollary 1 is proved in Appendix A with two mutually dual linear programs. The *primal* problem chooses a contract to fulfill some given objective subject to incentive compatibility.[13] Its *dual* problem has contracts as multipliers and chooses undetectable deviation plans. This motivates Definition 3 as a "backward-engineering" exercise: what minimal requirement on a monitoring technology yields multipliers on incentive constraints equal to zero (i.e., incentive constraints do not bind)?

When can secret (i.e., recommendation-contingent) contracts add value over and above standard ones? In other words, when can secret contracts approximately enforce more action profiles than standard ones? To motivate, consider an example.

[13] Although no budget constraints were imposed, we could have added *expected* budget balance,

$$\sum_{(i,a,s)} \xi_i(a, s) \;=\; 0,$$

but this constraint would not bind, since adding a constant to any ξ preserves its incentive properties.

Example 1. There are two publicly verifiable signals, $S = S_0 = \{x, y\}$, and two players, $I = \{1, 2\}$. Player 1 has two actions, $A_1 = \{U, D\}$, and player 2 has three actions, $A_2 = \{L, M, R\}$. The conditional probability system Pr is given below.

	L	M	R
U	1, 0	0, 1	1/2, 1/2
D	0, 1	1, 0	1/3, 2/3

$\Pr(U, R)$ clearly lies in the convex hull of $\Pr(U, L)$ and $\Pr(U, M)$. Intuitively, there is a mixed deviation (namely $\frac{1}{2}[L] + \frac{1}{2}[M]$, where $[\cdot]$ stands for Dirac measure) by player 2 such that the conditional probability over signals is indistinguishable from what it would be if he played R. In fact, a similar phenomenon takes place when player 1 plays D (this time with mixed deviation $\frac{2}{3}[L] + \frac{1}{3}[M]$) or indeed regardless of player 1's mixed strategy. It is therefore impossible to even approximately enforce R with transfers contingent only on signals if player 2 strictly prefers playing L and M, since there always exists a profitable deviation without any contractual losses.

However, Pr detects unilateral disobedience, so by Corollary 1 the profile (U, R) can be approximately enforced even if R is strictly dominated by both L and M. By correlating player 2's payment with player 1's recommendation, secret contracts can keep player 2 from knowing the proportion with which he ought to mix between L and M in order for his contractual payment to equal what he would obtain by playing R. This suggests how secret contracts can extract more information from a monitoring technology to provide incentives, even with publicly verifiable signals.

In general, the limited scope of standard contracts is characterized below. Given $\sigma \in \Delta(A)$, a monitoring technology Pr *detects unilateral disobedience* at σ (DUD-σ) if $\Pr(\sigma) \neq \Pr(\sigma, \alpha_i)$ for every player i and every disobedient deviation plan α_i.

Corollary 2. *Fix any monitoring technology* Pr*. Any team with any profile of utility functions can approximately enforce any correlated strategy with just signal-contingent contracts if and only if for every correlated strategy σ there is a sequence $\{\sigma^m\}$ of correlated strategies such that $\sigma^m \to \sigma$ and* Pr *satisfies DUD-σ^m for all $m \in \mathbb{N}$.*

With secret contracts, different correlated strategies may be used to detect different deviation plans, whereas with standard contracts, the same correlated strategy must detect all deviation plans by all players in order to characterize approximate enforcement. For instance, in Example 1 there is no sequence $\{\sigma^m\}$ of correlated strategies converging to $[(U, R)]$ with Pr satisfying DUD-σ^m for all m, yet DUD holds.

Let us relate DUD to the literature. If monitoring is publicly verifiable (i.e., S_i is a singleton for all $i \neq 0$), DUD reduces to the following *convex independence* (CI):

$$\forall(i, a_i), \qquad \Pr(a_i) \notin \text{conv}\{\Pr(b_i) : b_i \neq a_i\} \qquad \text{in } \mathbb{R}^{A_{-i}\times S},$$

where $\Pr(b_i) \in \mathbb{R}^{A_{-i}\times S}$ is given by $\Pr(b_i)(a_{-i}, s) = \Pr(s|a_{-i}, b_i)$ and "conv" stands for convex hull. This is substantially weaker than the following condition, call it *exact convex independence* (ECI), where $\Pr(a) \in \mathbb{R}^S$ is defined by $\Pr(a)(s) = \Pr(s|a)$:

$$\forall(i, a), \qquad \Pr(a) \notin \text{conv}\{\Pr(a_{-i}, b_i) : b_i \neq a_i\} \qquad \text{in } \mathbb{R}^S.$$

ECI means that signal probabilities conditional on any action profile change with every (mixed) unilateral deviation. In Section 4.2 it is shown that ECI is necessary and sufficient to *exactly* enforce any correlated strategy (without budget-balance). CI is substantially weaker than ECI, since CI is necessary and sufficient to *approximately* enforce any correlated strategy (Corollary 1). Formally, both CI and ECI require certain vectors to lie outside some convex hull. For CI, the vectors have dimension $A_{-i} \times S$, while for ECI they only have dimension S. Intuitively, CI requires that every deviation plan be detectable, allowing for different correlated strategies to detect different deviations, whereas ECI requires detectability with the same correlated strategy across all deviations and players.

Legros and Matsushima (1991) and Legros and Matthews (1993) provide conditions equivalent to ECI (but differently interpreted) and establish exact enforcement with standard contracts. In repeated games, and to prove folk theorems, *individual full rank* (IFR) has been prominent in the literature with public monitoring (Fudenberg et al., 1994). Formally, IFR (at some σ) means that $\Pr(\sigma) \notin \text{lin}\{\Pr(\sigma, b_i) : b_i \neq a_i\}$ for every i, where "lin" stands for linear span. Arguably, the spirit of IFR (and how it is used in practice) is to detect deviations away from some prescribed σ, i.e., ECI at σ: $\Pr(\sigma) \notin \text{conv}\{\Pr(\sigma, b_i) : b_i \neq a_i\}$ for every i. Clearly, IFR implies ECI, but not conversely. If $|S| < |A_i|$ for some i then this holds trivially, since IFR is impossible yet ECI is possible (e.g., all the points on a circle are convexly independent). This holds even with at least as many signals as actions (e.g., the vectors $(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$, $(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{1}{6}, 0, \frac{1}{3}, \frac{1}{2})$ and $(\frac{1}{3}, 0, \frac{1}{6}, \frac{1}{2})$ are convexly independent but linearly dependent). CI is also weaker than *local individual full rank* (LIFR) of d'Aspremont and Gérard-Varet (1998), which requires IFR at some σ, possibly different for each i.[14] This is true even in spirit, as Example 1 shows, since even "local ECI" fails there.

[14]For CI, every deviation plan can be detected with different correlated strategies for each plan, whereas LIFR uses the same correlated strategy across deviation plans, though not across players.

DUD is still weak even if monitoring is not verifiable. It is weaker than generalizations of IFR in Compte (1998), Kandori and Matsushima (1998), Obara (2005) and Tomala (2005)[15] used to prove folk theorems with private monitoring (and communication), as well as a condition by Obara (2006) in mechanism design (Lemma B.2).

We conclude this section by establishing genericity of DUD. Given a player with at least two observations, there must be at least as many action-signal pairs for others as for that player. Given a player without observations to report, genericity requires slightly fewer action-signal pairs for others: at least as many action-signal pairs for others with one signal omitted as actions for that player with one action omitted.

Theorem 2. *DUD is generic if and only if* $|A_i \times S_i| \leq |A_{-i} \times S_{-i}|$ *for every* i *such that* $|S_i| > 1$ *and* $|A_i| - 1 \leq |A_{-i}| \times (|S_{-i}| - 1)$ *for every* i *such that* $|S_i| = 1$.

Proof. For necessity, by Lemma B.2, DUD is implied by *convex independence* (CI):

$$\forall (i, a_i, s_i), \qquad \Pr(a_i, s_i) \notin \mathrm{conv}\{\Pr(b_i, t_i) : (b_i, t_i) \neq (a_i, s_i)\}.$$

In turn, CI is implied by *linear* independence, or full row rank, for all i, of the matrix with $|A_i \times S_i|$ rows, $|A_{-i} \times S_{-i}|$ columns and entries $\Pr(a_i, s_i)(a_{-i}, s_{-i}) = \Pr(s|a)$. Since the set of full rank matrices is generic, if $|S_i| > 1$ then this is satisfied generically when $|A_i \times S_i| \leq |A_{-i} \times S_{-i}|$. If $|S_i| = 1$, then adding with respect to s_{-i} for each a_{-i} yields column vectors equal to $(1, \ldots, 1) \in \mathbb{R}^{A_i}$. Therefore, there are $A_{-i} - 1$ linearly dependent columns. Eliminating them, it follows that genericity requires

$$|A_i| = |A_i \times S_i| \leq |A_{-i} \times S_{-i}| - (|A_{-i}| - 1) = 1 + |A_{-i}| \times (|S_{-i}| - 1).$$

This must hold for all i. Since the intersection of finitely many generic sets is generic, necessity follows. For a proof of sufficiency, see Appendix A. □

If $|S| = 1$ then DUD is generic only if $|A| = 1$. More interestingly, DUD is generic even if $|S| = 2$, as long as players have enough actions. Hence, a team may overcome incentive constraints (i.e., Corollary 1 holds) generically even if only one individual is able to make substantive observations and these observations are just a simple binary bit of information. If others' action spaces are large enough and their actions have generic effect on the bit's probability, this uniquely informed individual could still be controlled by testing him with unpredictable combinations of others' actions.[16]

[15] Here, detection is defined with respect to the same correlated strategy for each deviation plan.

[16] We thank an anonymous referee for urging us to emphasize this point.

3.2 Attribution

Definition 5 (Attribution). A deviation plan α_i for player i is *unattributable* if there exists a profile $\alpha_{-i} = (\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n)$ of deviation plans such that

$$\forall \sigma \in \Delta(A), \qquad \Pr(\sigma, \alpha_1) \;=\; \cdots \;=\; \Pr(\sigma, \alpha_i) \;=\; \cdots \;=\; \Pr(\sigma, \alpha_n).$$

Call α_i *attributable* if it is not unattributable, i.e., for every profile α_{-i} of deviation plans, there is a correlated strategy σ and a player j such that $\Pr(\sigma, \alpha_i) \neq \Pr(\sigma, \alpha_j)$.

Intuitively, a deviation plan is unattributable if there exists a profile of opponents' deviation plans such that every unilateral deviation would lead to the same expected report probabilities. Heuristically, after an unattributable unilateral deviation, even if the fact that someone deviated is detected, anyone could have been the culprit.

Definition 6 (IOP). A monitoring technology Pr *identifies obedient players* (IOP) if every disobedient deviation plan is attributable.

IOP is a stronger requirement on a monitoring technology than DUD. Indeed, DUD follows by replacing α_j above with honesty and obedience. IOP means that any profile of disobedient deviation plans that affects the probability of reported signals must do so in a way that is different for some players, since otherwise they would be unattributable. Conversely, if IOP fails then there exist disobedient deviation plans that change conditional probabilities in the same way for every player, so anyone could have disobeyed. Budget-balanced implementation must therefore fail, since players' incentives would "overlap." In other words, it would be impossible to punish some and reward others at the same time in order to provide adequate incentives. If all players must be punished or rewarded together, then budget balance must fail.

In comparison with Holmström (1982), who appointed a principal to play the role of budget-breaker, it will be seen that a team whose monitoring technology exhibits IOP can share that role internally. In some teams, this might be allocated stochastically, even leading to a secret principal (Section 2.2). Indeed, Holmström (1982)'s principal works because after a unilateral deviation, the principal (having no actions to take) can be identified as an obedient player, i.e., obedience is 'attributable' to the principal. Heuristically, IOP emphasizes rewarding the innocent rather than punishing the guilty. Furthermore, Corollary 3 below argues that the former perspective on incentives delivers informational economies relative to the latter.

Theorem 3. *A monitoring technology identifies obedient players if and only if it provides strict incentives with budget balance, i.e., there exists a probability weighted incentive scheme* $\xi : I \times A \times S \to \mathbb{R}$ *such that* $\sum_i \xi_i(a, s) = 0$ *for every* (a, s)*, and*

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

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

Proof. By the Alternative Theorem (Rockafellar, 1970, Theorem 22.2, p. 198), Pr fails to provide strict incentives with budget balance if and only if there exist vectors $\lambda \geq 0$ and $\eta \in \mathbb{R}^{A\times S}$ such that $\lambda_i(a_i, b_i, \rho_i) > 0$ for some (i, a_i, b_i, ρ_i) with $a_i \neq b_i$ and

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

where η is independent of i. Let $\Lambda = \max_{(i,a_i)} \sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) > 0$. For every player i, let α_i be the deviation plan defined pointwise by

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

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

Corollary 3. *A monitoring technology identifies obedient players if and only if any team with any profile of utility functions can approximately enforce any correlated strategy with budget balanced secret contracts.*

The proof of this result is almost identical to that of Corollary 1, therefore omitted. The only difference is that the primal includes (∗∗), yielding a slightly different dual.

In the context of *publicly verifiable monitoring*, IOP reduces to DUD together with

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

where $\mathbf{0}$ stands for the origin of $\mathbb{R}^{A\times S}$ and for every i, C_i (called the *cone* of player i) is the set of all vectors $\eta \in \mathbb{R}^{A\times S}$ such that for some deviation plan $\alpha_i : A_i \to \Delta(A_i)$,

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

Call this condition on $\{C_i\}$ *non-overlapping cones* (NOC). Fudenberg et al. (1994) impose a full rank condition for each pair of players at each action profile, implying that certain hyperplanes intersect only at the origin for *every pair* of players. On the other hand, NOC requires that certain cones intersect only at the origin for *all* players. Thus, it is possible that two players' cones overlap, i.e., their intersection is larger than just the origin. In general, NOC does not even require that there always be two players whose cones fail to overlap, in contrast with the compatibility condition of d'Aspremont and Gérard-Varet (1998), as Figure 1 below illustrates.[17]

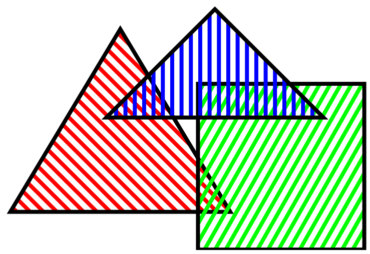

Figure 1: A cross-section of three non-overlapping cones in $\mathbb{R}^3$ (pointed at the origin behind the page) such that every pair of cones overlaps.

Upon a unilateral disobedience that changes probabilities by DUD, although it may be impossible to identify deviator(s), there must be someone to who could not have generated the statistical change. This way, IOP identifies obedient players. Budget balanced incentives are now possible, rewarding the obedient and punishing all others.

Just as for DUD, IOP can be translated to an equivalent condition with dual economic interpretation. The condition is PSI with budget balance, and its equivalence to IOP follows by the same argument as for DUD and PSI. Specifically for (publicly) verifiable monitoring, the fact that IOP can be decomposed into two separate conditions, DUD and NOC, provides useful insights, as shown next.

Definition 7. A verifiable monitoring technology Pr *clears every budget* (CEB) if given $K : A \times S \to \mathbb{R}$ there exists $\xi : I \times A \times S \to \mathbb{R}$ such that

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

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

The function $K(a, s)$ may be regarded as a budgetary surplus or deficit for each combination of recommended action and realized signal. CEB means that any level

[17]IOP is weaker than pairwise full rank, local pairwise full rank, and compatibility of d'Aspremont and Gérard-Varet (1998) also in the sense of approximate versus exact implementation (as well as information extraction), like DUD versus (local) individual full rank in Example 1.

of such budgetary surplus or deficit can be attained by a team without disrupting any incentive compatibility constraints. As it turns out, this is equivalent to NOC.

Proposition 1. *A verifiable monitoring technology has non-overlapping cones if and only if it clears every budget.*

This result further clarifies the relative roles of DUD and NOC. By Theorem 1, DUD characterizes approximate enforceability of any action profile a by secret contract. However, the team's budget may not be balanced ex post. NOC guarantees 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 approximately enforce a with ex post budget balance.[18]

Without verifiability, a decomposition of IOP into two separate parts does not emerge naturally. Indeed, it is not difficult to see that NOC plus DUD is sufficient but not necessary for IOP. To see this, notice there exist deviations, namely dishonest but obedient ones, that do not directly affect anyone's utility, and as such IOP allows them to remain unattributable (like DUD). With verifiability, every deviation may in principle affect players directly.

Example 2. Suppose there exists an individual i_0 such that A_{i_0} and S_{i_0} are both singleton sets. Here, DUD suffices for approximate implementability with ex post budget balance for this team, since player i_0 cannot be a deviator. She may become a "principal" and serve as "budget-breaker," much like a seller in an auction.

Example 3. Consider a team with two players ($I = \{1, 2\}$) and two publicly verifiable signals ($S = S_0 = \{x, y\}$). The players play the normal-form game (left) with public monitoring technology (right) below:

	w	s_2
m	$2, -1$	$-1, 0$
s_1	$3, -1$	$0, 0$

Utility Payoffs

	w	s_2
m	$p, 1-p$	$q, 1-q$
s_1	$1/2, 1/2$	$1/2, 1/2$

Signal Probabilities

Suppose that $q > p > 1/2$. First we will show that the "desirable" profile (s_1, w) cannot even be implemented approximately with standard (i.e., non-secret) contracts. With any standard contract, player 1 must be indifferent between monitoring and

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

shirking to approximate efficiency (it can be shown that player 2's randomization does not help). This implies that $1 = \frac{1}{4}(\zeta_1(x) - \zeta_1(y))$, where $\zeta_1(\omega)$ is the transfer to player 1 when $\omega \in \{x, y\}$ realizes. Budget balance requires $1 = \frac{1}{4}(\zeta_2(y) - \zeta_2(x))$. Since player 2's incentive constraint is $1 \leq \sigma\frac{1}{4}(\zeta_2(y) - \zeta_2(x))$, where σ denotes the probability that player 1 plays m, it follows that σ cannot be smaller than 1.

There exist budget-balanced secret contracts that approximately implement (s_1, w). Indeed, let player 1 play m with any probability $\sigma > 0$ and player 2 play w with probability 1. Let $\zeta : A \times S \to \mathbb{R}$ denote monetary transfers to player 1 from player 2, and fix $\zeta(a, s) = 0$ for all (a, s) except (m, w, x). That is, no money is transferred at all except when (m, w) is recommended and x realizes. Clearly, s_1 and s_2 are incentive compatible when recommended. The remaining incentive constraints simplify to:

$$\begin{aligned} m: \qquad & 1 + \frac{\sigma(m,w)}{\sigma(m)}(\tfrac{1}{2} - p)\zeta(m,w,x) \ \leq \ 0 \\ w: \qquad & 1 + \frac{\sigma(m,w)}{\sigma(w)}(p - q)\zeta(m,w,x) \ \leq \ 0 \end{aligned}$$

These two inequalities can clearly be satisfied by taking $\zeta(m, w, x)$ large enough. It is not difficult to check that IOP is satisfied (hence also DUD) if and only if $p \neq q$ and $(p-1/2)(q-1/2) > 0$. Thus, Robinson and Friday (Section 2.1) cannot approximately enforce (s_1, w) with budget balance.

Example 4. Without verifiability ($S = S_1 = \{x, y\}$) IOP fails, but the same condition suffices to approximately enforce (s_1, w) with budget balance. However, not everything is approximately enforceable. See Section 4.2 for conditions on a monitoring technology to approximately enforce a given action profile.

Finally, we establish genericity.

Proposition 2. *IOP is generic if also* $\left|A_{-\{i,j\}} \times S_{-\{i,j\}}\right| \geq |A_i \times S_i| + |A_j \times S_j| - 1$ *for some* $i, j \in I$.

Proof. □

4 Discussion

This section makes four comments. Firstly, it fills an important gap in the interpretation of Theorems 1 and 3. Secondly, it reconciles our main results with the literature by applying the duality of our model to the case of fixed action profiles and utility functions. Thirdly, environmental complications such as limited liability and individual rationality are examined, where standard results generalize to our setting easily, such as that only total liability matters to a team or that individual rationality is not a binding constraint. We end the section by arguing that DUD and IOP, as well as similar variants, are generic in relatively low dimensional spaces.

4.1 Exact versus Approximate Enforcement

A correlated strategy σ is (exactly) *implementable* if there is a scheme ζ such that

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

In Section ??, approximate implementability is defined in terms of linear inequalities: σ is approximately implementable if a ξ exists such that (σ, ξ) satisfies $(*)$. To justify, it must be shown that (σ, ξ) is *approachable*: there is a sequence $\{(\sigma^m, \zeta^m)\}$ such that (σ^m, ζ^m) satisfies $(***)$ for every m, $\sigma^m \to \sigma$, and $\sigma^m\zeta^m \to \xi$. The next result proves this under DUD and IOP. In addition, IOP implies every action profile is approachable with contracts that are budget balanced "along the way," not just asymptotically.

Proposition 3. *Pr satisfies DUD (IOP) only if every completely mixed correlated strategy is implementable (with budget balance). Hence, DUD (IOP) implies that every contract satisfying $(*)$ (and $(**)$) is approachable (with budget balance).*

When DUD or IOP fails, the "closure" of $(*\,*\,*)$ does not necessarily equal $(*)$. To illustrate, consider the following variation of Robinson and Friday (Section 2.1):

	work	shirk	rest
monitor	$2,-1$	$-1,0$	$-1,0$
shirk	$3,-1$	$0,0$	$0,-1$

Utility Payoffs

	work	shirk	rest
monitor	$1,0$	$0,1$	$1,0$
shirk	$1/2,1/2$	$1/2,1/2$	$1/2,1/2$

Signal Probabilities

Assume the signal is public. The profile (shirk,work) is approximately implementable with transfers ξ given by $\xi_F(g|\text{monitor,work}) = 1$ and $\xi_i(a, s) = 0$ for other (i, a, s). However, since rest is indistinguishable from work and rest weakly dominates work, no contract can dissuade Friday from resting. Hence, (shirk,work) is not approachable. Generalizing Proposition 3 involves iterated elimination of weakly dominated *indistinguishable* strategies in the spirit of Myerson's (1997) dual reduction; details are left for another paper. (But Theorem 4 below provides a partial generalization.)

4.2 Fixed Action Profiles and Utility Functions

A characterization of implementable action profiles also follows. We focus on budget balanced implementation (without proof, since it is just like that of Theorem 3); the unbalanced case—being similar—is omitted.

A mixed deviation $\alpha_i \in \Delta(\Theta_i)$ for player i is *unattributable* at $a \in A$ if there is a profile of mixed deviations α such that $\Pr(\alpha_i, a) = \Pr(\alpha_j, a)$ for every j, where $\Pr(\alpha_i, a)(s) = \sum_{\theta_i} \alpha_i(\theta_i) \Pr(s|\theta_i, a_{-i})$ for every s; otherwise it is *attributable* at a. Say Pr *identifies obedient players* at a (IOP-a) if every mixed deviation α_i with $\alpha_i(\rho_i, b_i) > 0$ for some $b_i \neq a_i$ is attributable at a.

Proposition 4. *A monitoring technology identifies obedient players at an action profile a if and only if any team with any profile of utility functions can exactly implement a with budget balanced secret contracts.*

With verifiable monitoring, IOP-a can be decomposed into two conditions. The first is *exact convex independence* at a (ECI-a), which means that the requirement for ECI from Section 3.1 holds at a. For the second, let $C_i(a)$ be the *cone* of player i at a, i.e., the set of all vectors $\eta \in \mathbb{R}^S$ such that for some mixed deviation α_i,

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

A verifiable monitoring technology Pr has *non-overlapping cones* at a (NOC-a) if

$$\bigcap_{i \in I} C_i(a) \;=\; \{\mathbf{0}\}.$$

IOP-a is equivalent to ECI-a and NOC-a. It generalizes the famous *pairwise full rank* condition of Fudenberg et al. (1994), and implies (but is not implied by) for all $i \neq j$,

$$C_i(a) \cap C_j(a) \;=\; \{\mathbf{0}\}.$$

Intuitively, i's and j's deviations can be statistically distinguished at a. On the other hand, NOC-a allows some players' cones to overlap. Naturally, this is weaker than pairwise full rank at a, and generally even weaker than $C_i(a) \cap C_j(a) = \{\mathbf{0}\}$ for some i, j. Intuitively, NOC-a requires that some player can be identified as obedient at a.[19]

It is possible to partially generalize Proposition 3 by fixing utility functions. To this end, a deviation plan α_i for i is v_i-*detectable* if $\Pr(\sigma) = \Pr(\alpha_i, \sigma)$ for every correlated strategy σ implies $v_i(\alpha_i, \sigma) \leq v_i(\sigma)$ for every σ, where $v_i(\sigma) = \sum_a v_i(a)\sigma(a)$ and $v_i(\sigma) = \sum_{(a,\theta_i)} v_i(b_i, a_{-i})\alpha_i(\theta_i|a_i)\sigma(a)$. Pr v-*detects unilateral disobedience* (v-DUD) if every disobedient deviation plan α_i of any player i is v_i-detectable. Similarly, call α_i v-*attributable* if existence of a profile α of deviation plans with $\Pr(\alpha_i, \sigma) = \Pr(\alpha_j, \sigma)$ for every σ and every j implies that $\sum_i v_i(\alpha_i, \sigma) - v_i(\sigma) \leq 0$. Pr v-*identifies obedient players* (v-IOP) if every disobedient deviation plan is v-attributable.

The next result follows immediately from the duality of Theorem 1 and Proposition 3, so its proof is omitted. It could also be extended to describe exact implementability in line with Proposition 4 after suitably amending v-DUD/v-IOP; details are left to the reader.

Theorem 4. *A monitoring technology exhibits v-DUD (v-IOP) if and only if any action profile is approximately implementable with (budget balanced) secret contracts. Furthermore, Proposition 3 still holds with v-DUD (v-IOP) replacing DUD (IOP).*

4.3 Participation and Liability

In this subsection we will use duality to study teams subject to liquidity constraints. One such constraint is *limited liability*, where an individual's transfers are bounded below. This can be taken into account by adding $\zeta_i(a, s) \geq \ell_i$ or $\xi_i(a, s) \geq \sigma(a)\ell_i$ to the metering problem, where ℓ_i is an exogenous parameter representing player i's *liability*. Let $\ell = (\ell_1, \ldots, \ell_n)$ be the profile of liabilities faced by a team. A team's *total liability* is defined by $\widehat{\ell} = \sum_i \ell_i$. By a simple duality and without restrictions on a team's monitoring technology, we can generalize to our setting Theorem 5 of Legros and Matsushima (1991) and Theorem 4 of Legros and Matthews (1993).

[19] Restricted to public monitoring, Proposition 4 is equivalent to Proposition 3 in Legros and Matsushima (1991). Similar results are also in Lemma 1 of d'Aspremont and Gérard-Varet (1998), but our decomposition and interpretation are new. Indeed, IOP-a is weaker than their compatibility, which requires pairwise full rank for some pair of players.

Proposition 5. *Only total liability affects a team's (approximately) implementable action profiles (with and without budget balance).*

It is possible that the team faces double-sided limited liability, which may be captured by adding a version of the following constraints to the metering problem:

$$\forall (i, a, s), \qquad -\sigma(a)\ell_i \;\leq\; \xi_i(a, s) \;\leq\; \sigma(a)\ell_i,$$

for some $\ell_i \geq 0$. These constraints lead to an alternative, linear way of requiring that ξ be adapted to σ (i.e., $\xi_i(a, s) = 0$ whenever $\sigma(a) = 0$).

Individual rationality is also amenable to our study of incentives. Without budget balance, since players can be paid lump sums to become indifferent between belonging to the team and forsaking it, individual rationality constraints cannot bind. Hence, suppose the team's budget must be balanced ex post. As a normalization, assume that $\sum_i v_i(a) \geq 0$ for all $a \in A$. Participation constraints may be incorporated as:

$$\forall i \in I, \qquad \sum_{a \in A} \sigma(a) v_i(a) + \sum_{s \in S} \xi_i(a, s) \Pr(s|a) \;\geq\; 0.$$

Proposition 6. *Participation is not a binding constraint if $\sum_i v_i(a) \geq 0$ for all a.*

5 Conclusion

In this paper we have explored possible ways in which secret contracts may help organizations, with particular emphasis on the question of monitoring a monitor and maintaining budget balance. Formally, we have used duality systematically to make general statements about a team's contractual scope. We have exploited this duality to consider teams with infinitely many actions and signals, with fruitful applications such as a subdifferential characterization equilibrium payoffs. Below, we conclude this paper with some comments to connect the paper with the (mechanism design and implementation) literature, discuss weaknesses (collusion), and further research.

5.1 Abstract Mechanisms in Concrete Contracts

We build a bridge between abstract mechanism design and concrete contract theory in this paper. Some of the mechanism design literature has focused on surplus extraction in environments with adverse selection. Thus, Cremer and McLean (1988) argued that if individuals have "correlated types" then their surplus may be extracted.[20] On the other hand, they do not explain the source of such correlation. Secret contracts provide an explanation for the emergence of correlated types.

As part of a team's economic organization, it may be beneficial for private information to be allocated differently in order to provide the right incentives. As has been argued here, this is true even if the team starts without informational asymmetry. In a sense, correlated types emerge endogenously, and as such there are incidental similarities between this paper and the mechanism design literature even if conceptually there are important differences. For instance, the essence of secret contracts is lost in the abstraction of mechanism design because it so reduced. With moral hazard, our identifiability conditions apparently lend themselves easily to interpretation.

Nonetheless, a hybrid exercise where players begin with some private information and face an additional metering problem is amenable to the techniques developed here. Initial results are promising (Rahman, 2005b, Ch. 5); details are for another paper.

5.2 Secrets and Verifiable Recommendations

Secret contracts rely on making payments contingent on verifiable recommendations. Even if the mediator's messages are unverifiable, it may still be possible for players to verifiably reveal their recommendations. Player i's reporting strategy would then involve announcing a recommended action and a private signal. Incentive constraints would be only slightly different.

Kandori (2003) used similar schemes in Nash equilibrium for repeated games with public monitoring, by having players mix independently and transfers depend on reported realizations of mixed strategies. Our framework is more general because we study private monitoring with communication in correlated equilibrium. Moreover, we do not require pairwise conditions on the monitoring technology for a folk theorem.

As illustrated by Robinson and Friday in Section 2.1, secret contracts provide an intuitive organizational design. If recommendations were not verifiable, then in order to approximate efficiency Friday would need to report whether or not he worked,

[20] d'Aspremont et al. (2004) extend this result to include budget balance. The additional constraint of individual rationality is studied by Kosenok and Severinov (2004).

which broadly interpreted provides a different answer to the question of monitoring the monitor: have two monitors monitoring each other. We purposely avoided this.

5.3 Usual Problems with Collusion

[Discuss the model with multilateral deviations.]

A notable weakness of secret contracts is not being collusion-proof. To illustrate, in our leading example (Section 2.1) Robinson and Friday could communicate to break down the incentives that secrets tried to provide. However, this problem is neither inherent to secrets nor widespread to all teams. Example 3 describes when Robinson and Friday can approximate efficiency with budget balance, for which they require secrets. There, contracts are naturally robust to collusion, since budget balance implies that Friday's gain is Robinson's loss.

Collusion is a problem for secret contracts inasmuch as it is a problem for contracts in general. For instance, the transfer schemes of Cremer and McLean (1988) are not generally collusion-proof for similar reasons. In any case, although there may be partial solutions to the problem of collusion with secret contracts in the spirit of, say, Che and Kim (2006), the main purpose of this paper is to introduce secret contracts. Thus, analysis of collusion is postponed for the future. Meanwhile, the scheme below weakly dissuades extra-contractual communication between Robinson and Friday.

	(monitor,work)	(monitor,shirk)	(shirk,work)	(shirk,shirk)
g	$1/\mu, 1/\sigma$	$0, 1/\sigma$	$1/2\mu, 0$	$0, 1/2(1-\sigma)$
b	$0, 0$	$1/(1-\mu), 0$	$0, 1/(1-\sigma)$	$1/2(1-\mu), 1/2(1-\sigma)$

A Proofs

Corollary 1. Consider the following linear program, called the *primal*.

$$V_f(v) := \sup_{\sigma \geq 0, \xi} \sum_{a \in A} f(a)\sigma(a) \quad \text{s.t.} \quad \sum_{a \in A} \sigma(a) = 1,$$

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

The *dual* is given below. By FTLP, the value of the dual equals that of the primal.

$$
\begin{aligned}
V_f(v) \;&=\; \inf_{\lambda \geq 0, \kappa} \quad \kappa \quad \text{s.t.} \\
\forall a \in A, \qquad \kappa \;&\geq\; f(a) - \sum_{(i,\theta_i)} \lambda_i(a_i,\theta_i)(v_i(b_i,a_{-i}) - v_i(a)) \\
\forall (i,a,s), \qquad \sum_{\theta_i \in \Theta_i} \lambda_i(a_i,\theta_i)&(\Pr(s|a) - \Pr(s|\theta_i,a_{-i})) \;=\; 0
\end{aligned}
$$

Clearly, λ is feasible if and only if its probabilistic normalization is undetectable. We will show that DUD is equivalent to $V_f(v) = \max\{f(a) : a \in A\}$ for all f. If Pr satisfies DUD then by the second family of dual constraints, any feasible $\lambda \neq 0$ must have $\lambda_i(a_i,\theta_i) > 0$ only if $a_i = b_i$. Hence, the first family of dual constraints becomes $\kappa \geq f(a)$ for all a. Minimizing κ subject to them yields $\max\{f(a) : a \in A\}$ for any f and v, proving sufficiency. For necessity, if DUD fails there is $\lambda \geq 0$ with

$$
\sum_{\theta_i \in \Theta_i} \lambda_i(a_i,\theta_i)(\Pr(s|a) - \Pr(s|\theta_i,a_{-i})) \;=\; 0
$$

for all (i,a,s) and $\lambda_j(\widehat{a}_j,\widehat{\theta}_j) > 0$ for some $(j,\widehat{a}_j,\widehat{\theta}_j)$ with $\widehat{b}_j \neq \widehat{a}_j$. Let $f = \mathbf{1}_{\widehat{a}_j}$ and choose v as follows. For any a_{-j}, the utility to each player depending on whether or not j plays $\widehat{a}_j$ is given by (first is j then anyone else):

a_j	$\widehat{a}_j$
1, 0	0, 2

Given a with $a_j \neq \widehat{a}_j$, the first dual constraint becomes $0 + \sum_{\rho_j} \lambda(a_j,\widehat{a}_j,\rho_j) \leq \kappa$. This can be made smaller than 1 by multiplying λ by a sufficiently small positive number. At $\widehat{a}_j$, the constraint becomes $1 - \sum_{\theta_j} \lambda_j(\widehat{a}_j,\theta_j) \leq \kappa$. Since $\sum \lambda > 0$, there is a feasible dual solution with $\kappa < 1 = \max\{f(a)\}$, as required. □

Proposition 1. Consider the following primal problem: Find a feasible ξ to solve

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

The dual of this problem is given by

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

If CEB is satisfied, then the value of the primal equals 0 for any $K : A \times S \to \mathbb{R}$. By FTLP, 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 NOC is satisfied. For necessity, if NOC is satisfied then the value of the dual is always 0 for any $K : A \times S \to \mathbb{R}$. By FTLP, 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 the primal constraints, and CEB is satisfied. □

Proposition 3. For $B \subset A$, the B-cone generated by unidentifiable deviation profiles is

$$\mathcal{K}(B) \ := \ \{\lambda \geq 0 : \forall i \in I, a \in B, s \in S, \sum_{\theta_i \in \Theta_i} \lambda_i(a_i, \theta_i)(\Pr(s|a) - \Pr(s|\theta_i, a_{-i})) = 0\}.$$

By the Alternative Theorem (Rockafellar, 1970, p. 198), a given σ is implementable, i.e., there exists ζ to solve $(***)$, if and only if the following dual inequalities are satisfied:

$$\forall \lambda \in \mathcal{K}(\text{supp } \sigma), \qquad \sum_{(i,a_i,\theta_i)} \lambda_i(a_i, \theta_i) \sum_{a_{-i}} \sigma(a)(v_i(b_i, a_{-i}) - v_i(a)) \ \leq \ 0.$$

In contrast, approximate implementability of σ as in Definition 1 is equivalent to the smaller system of inequalities indexed instead by $\lambda \in \mathcal{K}(A) \subset K(\text{supp } \sigma)$. (Hence, exact implementability implies approximate.) Now, if σ is completely mixed then $\sigma(a) > 0$ for all a, so $\mathcal{K}(\text{supp } \sigma) = \mathcal{K}(A)$. By DUD, $\mathcal{K}(A)$ consists of all $\lambda \geq 0$ with $\lambda_i(a_i, \theta_i) > 0$ implying $a_i = b_i$, where $\theta_i = (\rho_i, b_i)$. Therefore, $\sum_{(i,a_i,\theta_i)} \lambda_i(a_i, \theta_i) \sum_{a_{-i}} \sigma(a)(v_i(b_i, a_{-i}) - v_i(a)) = 0$, and implementability follows. For IOP, replacing $\mathcal{K}(B)$ with

$$\mathcal{K}_0(B) := \{\lambda \geq 0 : \forall i \in I, a \in B, s \in S, \sum_{\theta_i \in \Theta_i} \lambda_i(a_i, \theta_i)(\Pr(s|a) - \Pr(s|\theta_i, a_{-i})) = \eta(a, s)\}$$

leads to the corresponding result by an almost identical argument.

Clearly, the closure of the space of contracts satisfying $(***)$ (and $(**)$) is contained in the space of contracts satisfying $(*)$ (and $(**)$), so it remains only to show the converse containment. To this end, pick any (σ, ξ) satisfying $(*)$ (and $(**)$). By the previous argument, the uniformly distributed correlated strategy with full support $\sigma^0 = (1/|A|, \ldots, 1/|A|)$ is implementable (with budget balance). For any sequence of positive probabilities $\{p_m\}$ decreasing to 0, consider the sequence of contracts $\{(\sigma^m, \zeta^m)\}$ defined for every (i, a, s) by $\sigma^m(a) = p_m \sigma^0(a) + (1 - p_m)\sigma(a)$ and $\zeta_i^m(a, s) = p_m \zeta_i^0(a, s) + (1 - p_m)\xi_i(a, s)/\sigma^m(a)$. This sequence of contracts converges to (σ, ξ) and satisfies $(***)$ (as well as $(**)$) for all m. □

Proposition 5. We just prove the result with budget balance; the rest follows similarly. The dual of the metering problem of maximizing $\sum_a f(a)\sigma(a)$ subject to limited liability, approximate implementability, and budget balance is

$$V_f(v, \ell) \ = \ \inf_{\lambda, \mu \geq 0, \eta, \kappa} \ \kappa \quad \text{s.t.}$$

$$\forall a \in A, \qquad \kappa \ \geq \ f(a) - \sum_{(i,\theta_i)} \lambda_i(a_i, \theta_i)(v_i(b_i, a_{-i}) - v_i(a)) - \sum_{(i,s)} \mu_i(a, s)\ell_i,$$

$$\forall (i, a, s), \qquad \sum_{\theta_i \in \Theta_i} \lambda_i(a_i, \theta_i)(\Pr(s|a) - \Pr(s|\theta_i, a_{-i})) + \mu_i(a, s) \ = \ \eta(a, s),$$

where $\mu_i(a, s)$ is a multiplier on the liquidity constraint for player i at (a, s). Adding the last family of equations with respect to s implies $\sum_s q_i(a, s) = \sum_s \eta(a, s)$ for every i. Therefore,

$$\sum_{(i,s)} \mu_i(a, s)\ell_i \ = \ \sum_{(i,s)} \eta(a, s)\ell_i \ = \ \sum_{s \in S} \eta(a, s)\widehat{\ell},$$

where $\widehat{\ell} = \sum_i \ell_i$, so we may eliminate $\mu_i(a,s)$ from the dual problem as follows:

$$V_f(v,\ell) \;=\; \inf_{\lambda,\eta,\kappa} \quad \kappa \quad \text{s.t.}$$

$$\forall a \in A, \qquad \kappa \;\geq\; f(a) - \sum_{(i,\theta_i)} \lambda_i(a_i,\theta_i)(v_i(b_i,a_{-i}) - v_i(a)) - \sum_{s\in S} \eta(a,s)\widehat{\ell}$$

$$\forall (i,a,s), \qquad \sum_{\theta_i\in\Theta_i} \lambda_i(a_i,\theta_i)(\Pr(s|a) - \Pr(s|\theta_i,a_{-i})) \;\leq\; \eta(a,s).$$

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

Proposition 6. The dual of the metering problem subject to participation is:

$$V_f(v) \;=\; \inf_{\lambda,\pi\geq 0,\kappa,\eta} \quad \kappa \quad \text{s.t.}$$

$$\forall a \in A, \qquad \kappa \;\geq\; f(a) - \sum_{(i,\theta_i)} \lambda_i(a_i,\theta_i)(v_i(b_i,a_{-i}) - v_i(a)) + \sum_{i\in I} \pi_i v_i(a)$$

$$\forall (i,a,s), \qquad \pi_i \Pr(s|a) + \sum_{\theta_i\in\Theta_i} \lambda_i(a_i,\theta_i)(\Pr(s|a) - \Pr(s|\theta_i,a_{-i})) \;=\; \eta(a,s)$$

where π_i is a multiplier for player i's participation constraint. Adding the second family of 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 all 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 minimized by $\pi = 0$. □

Sufficiency in Theorem 2. For sufficiency, suppose firstly that $|S_i| > 1$ yet $|A_i \times S_i| > |A_{-i} \times S_{-i}|$ for some i. For a fixed (a_i^0, s_i^0), let $\Pr(a_i^0, s_i^0) = (1/|S|, \ldots, 1/|S|)$. For all other (a_i, s_i),

[A proof of genericity via convex independence: if there are more points than dimensions (i.e., more rows than columns) then have all the vertices be one of the points. Then have another point (there's at least one left over by assumption) be the equally weighted average of all the vertices. Any remaining points can go anywhere. Clearly there is convex dependence. A small perturbation of all the points preserves convex dependence, so there is an open set of monitoring technologies with convex dependence. But an open set has positive Lebesgue measure. Therefore, SCI is generic if $|S| > 1$ and there are at least as many columns as rows.]

□

Corollary 2. By the Alternative Theorem (Rockafellar, 1970, Theorem 22.2, p. 198), for any correlated strategy σ, Pr satisfies DUD-σ if and only if there exists a signal contingent transfer scheme $\zeta : I \times S \to \mathbb{R}$ such that

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

with a strict inequality whenever $a_i \neq b_i$. Therefore, by scaling ζ appropriately, any deviation gains can be outweighed by contractual losses. The result now follows. □

B Lemmata

Lemma B.1. *DUD is equivalent to the following condition: If $\lambda \geq 0$ satisfies*

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

then $\lambda_i(b_i, t_i|a_i, s_i) = 0$ whenever $a_i \neq b_i$.

Proof.

□

Lemma B.2. *CI implies DUD.*

Proof.

□

References

Alchian, A. and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777–795. 1, 2

Aoyagi, M. (2005): "Collusion Through Mediated Communication in Repeated Games with Imperfect Private Monitoring," *Economic Theory*, 25, 455–475. 4

Che, Y.-K. and J. Kim (2006): "Robustly Collusion-Proof Implementation," *Econometrica*, 74, 1063–1107. 27

Compte, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66, 597–626. 4, 16

Cremer, J. and R. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257. 26, 27

d'Aspremont, C., J. Cremer, and L.-A. Gérard-Varet (2004): "Balanced Bayesian Mechanisms," *Journal of Economic Theory*, 115, 385–396. 20, 26

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. 3, 4, 15, 19, 24

Fudenberg, D., D. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997–1039. 4, 15, 18, 23

Holmström, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340. 2, 6, 7, 17

Kandori, M. (2003): "Randomization, Communication, and Efficiency in Repeated Games with Imperfect Public Monitoring," *Econometrica*, 71, 345–353. 4, 26

Kandori, M. and H. Matsushima (1998): "Private Observation, Communication, and Collusion," *Econometrica*, 66, 627–652. 4, 16

Kosenok, G. and S. Severinov (2004): "Individually Rational, Balanced-Budget Bayesian Mechanisms and the Informed Principal Problem," Mimeo. 26

Legros, P. and H. Matsushima (1991): "Efficiency in Partnerships," *Journal of Economic Theory*, 55, 296–322. 4, 15, 24

Legros, P. and S. Matthews (1993): "Efficient and Nearly Efficient Partnerships," *Review of Economic Studies*, 60, 599–611. 3, 15, 24

Myerson, R. (1986): "Multistage Games with Communication," *Econometrica*, 54, 323–358. 2

——— (1997): "Dual Reduction and Elementary Games," *Games and Economic Behavior*, 21, 183–202. 23

Nau, R. F. and K. F. McCardle (1990): "Coherent Behavior in Noncooperative Games," *Journal of Economic Theory*, 50, 424–444. 3

Obara, I. (2005): "Folk Theorem with Communication," Mimeo. 4, 16

——— (2006): "The Full Surplus Extraction Theorem with Hidden Actions," Mimeo. 16

Obara, I. and D. Rahman (2006): "Approximate Implementability with Ex Post Budget Balance," Mimeo. -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. 3, 7

Rahman, D. (2005a): "Optimum Contracts with Public and Private Monitoring," Mimeo. -1, 3

——— (2005b): "Team Formation and Organization," Ph.D. dissertation, UCLA. 26

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press. 13, 18, 29, 30

Strausz, R. (1997): "Delegation of Monitoring in a Principal-Agent Relationship," *Review of Economic Studies*, 64, 337–357. 2

Tomala, T. (2005): "On Subgame-Perfect Communication Equilibria in Repeated Games with Imperfect Monitoring," Mimeo. 4, 16